CORC Technical Report TR-2005-06
On two-stage convex chance constrained problems

E. Erdoğan * G. Iyengar †

November 3, 2005

Abstract

In this paper we develop approximation algorithms for two-stage convex chance constrained problems. Nemirovski and Shapiro [18] formulated this class of problems and proposed an ellipsoid-like iterative algorithm for the special case where the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine. We show that this algorithm extends to bi-convex $\mathbf{f}(\mathbf{x}, \mathbf{h})$ in a fairly straightforward fashion. The complexity of the solution algorithm as well as the quality of its output are functions of the radius r of the largest Euclidean ball that can be inscribed in the polytope defined by a random set of linear inequalities generated by the algorithm [18]. Since the polytope determining r is random, computing r is difficult. Yet, the solution algorithm requires r as an input. In this paper we provide some guidance for selecting r. We show that the largest value of r is determined by the degree of robust feasibility of the two-stage chance constrained problem – the more robust the problem, the higher one can set the parameter r.

Next, we formulate ambiguous two-stage chance constrained problems. In this formulation, the random variables defining the chance constraint are known to have a fixed distribution; however, the decision maker is only able to estimate this distribution to within some error. We construct an algorithm that solves the ambiguous two-stage chance constrained problem when the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine and the extreme points of a certain "dual" polytope are known explicitly.

1 Introduction

The simplest model for a convex chance constrained problem is as follows.

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X}_\epsilon(\mathbb{Q}) = \Big\{\mathbf{y} \in \mathcal{X} \mid \mathbb{Q}(\mathbf{H} : f(\mathbf{y}, \mathbf{H}) > 0) \leq \epsilon\Big\}, \end{aligned} \tag{1}$$

*IEOR Department, Columbia University, New York, New York 10027. Email: `ee168@columbia.edu`. Research partially supported by NSF grants CCR-00-09972, DMS-01-04282 and ONR grant N000140310514.

†IEOR Department, Columbia University, New York, New York 10027. Email: `gi10@columbia.edu`. Research partially supported by NSF grants CCR-00-09972, DMS-01-04282 and ONR grant N000140310514.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **03 NOV 2005**	2. REPORT TYPE	3. DATES COVERED **00-00-2005 to 00-00-2005**
4. TITLE AND SUBTITLE **On two-stage convex chance constrained problems**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Columbia University ,Department of Computer Science,New York,NY,10027**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In this paper we develop approximation algorithms for two-stage convex chance constrained problems. Nemirovski and Shapiro [18] formulated this class of problems and proposed an ellipsoid-like iterative algorithm for the special case where the impact function f (x, h) is bi-affine. We show that this algorithm extends to bi-convex f (x, h) in a fairly straightforward fashion. The complexity of the solution algorithm as well as the quality of its output are functions of the radius r of the largest Euclidean ball that can be inscribed in the polytope defined by a random set of linear inequalities generated by the algorithm [18]. Since the polytope determining r is random, computing r is difficult. Yet, the solution algorithm requires r as an input. In this paper we provide some guidance for selecting r. We show that the largest value of r is determined by the degree of robust feasibility of the two-stage chance constrained problem - the more robust the problem, the higher one can set the parameter r. Next, we formulate ambiguous two-stage chance constrained problems. In this formulation the random variables defining the chance constraint are known to have a fixed distribution; however, the decision maker is only able to estimate this distribution to within some error. We construct an algorithm that solves the ambiguous two-stage chance constrained problem when the impact function f (x, h) is bi-affine and the extreme points of a certain "dual" polytope are known explicitly.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**25**	

where $\mathbf{x} \in \mathbf{R}^n$ is the decision vector, the uncertain parameter $\mathbf{H} \in \mathbf{R}^m$ is distributed according to the *known* distribution $\mathbb{Q}$, $\mathcal{X} \subseteq \mathbf{R}^n$ is a convex set, and $f(\cdot, \mathbf{H}) : \mathbf{R}^n \mapsto \mathbf{R}$ is convex in $\mathbf{x}$ for each fixed $\mathbf{H}$. Since a collection of convex constraints $f_i(\mathbf{x}, \mathbf{H}) \leq 0$, $i = 1, \ldots, p$, can be formulated as a single convex constraint $f(\mathbf{x}, \mathbf{H}) = \max_{1 \leq i \leq p}\{f_i(\mathbf{x}, \mathbf{H})\} \leq 0$, restricting the range of constraint function $f(\cdot, \mathbf{H})$ to $\mathbf{R}$ does not represent any loss of generality. Also, by introducing a new variable if necessary, one can assume that the objective function is linear and independent of the uncertain parameter.

Chance constrained problems are computationally very difficult to solve. Except for a very restricted class of measures, evaluating $\mathbb{Q}(\mathbf{H} : f(\mathbf{x}, \mathbf{H}) > 0)$ involves numerically computing a multidimensional integral that becomes difficult as the number of parameters grows. Moreover, even if the function $f(\mathbf{x}, \mathbf{H})$ is linear in $\mathbf{x}$, the feasible set $\mathcal{X}_\epsilon(\mathbb{Q})$ of (1) is, in general, not convex. General chance constrained problems (i.e., $f(\mathbf{x}, \mathbf{H})$ is possibly non-convex) have a very rich literature. See [19, 22] for a survey of solution methods.

Recently there has been a renewed interest in the special case of convex chance constrained problems. One approximates the convex chance constrained problem (1) by the sampled optimization problem

$$\begin{aligned} \min \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{Y}[\mathbf{H}_{1,N}] = \{\mathbf{y} \in \mathcal{X} \mid f(\mathbf{y}, \mathbf{H}_i) \leq 0, i = 1, ..., N\}, \end{aligned} \tag{2}$$

where $\mathbf{H}_i$, $i = 1, \ldots, N$, are N IID samples from $\mathbb{Q}$. de Farias and Van Roy [9] use results from Learning theory [1, 15, 27] to show that for the special case of linear constraints a sample size of

$$N \geq \frac{4n}{\epsilon} \log\left(\frac{12}{\epsilon}\right) + \frac{4}{\epsilon} \log\left(\frac{2}{\delta}\right)$$

ensures that the set of decision vectors feasible for the sampled problem (2) is contained in $\mathcal{X}_\epsilon(\mathbb{Q})$ with a probability at least $1 - \delta$. Erdoğan and Iyengar [11] show a similar bound for general convex constraints with the constant n replaced by the Vapnik-Chervonenkis (VC) dimension d_f of the constraint function. Calafiore and Campi [7, 8] show that when

$$N \geq \frac{2n}{\epsilon} \ln\left(\frac{2}{\epsilon}\right) + \frac{2}{\epsilon} \ln\left(\frac{1}{\delta}\right) + 2n,$$

the optimal solution of the sampled problem (2) is feasible for (1) with a probability at least $1 - \delta$. This bound is particularly relevant since the VC dimension d_f of a constraint can be orders of magnitude larger than the problem dimension n. Note that these results only provide *upper* bounds for the number of samples, i.e. only a sufficient condition. Thus, a natural question of the quality or "tightness" of the approximation arises. Recently, Nemirovski and Shapiro [18, 17] established logarithmically separated upper and lower bounds on the number of samples required to approximate a convex chance constrained problem when the measure $\mathbb{Q}$ has a certain concentration-of-measure property.

The chance constrained problem assumes that the distribution $\mathbb{Q}$ of the random parameter $\mathbf{H}$ is known and fixed. In practice, however, the distribution $\mathbb{Q}$ is only specified with some error. Erdoğan and Iyengar [11] model this ambiguity in the measure by assuming that the measure $\mathbb{Q}$ is only known to belong to the set

$$\mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\},$$

where $\rho_p(\cdot,\cdot)$ denotes the Prohorov metric (see § 4). Given this description of the information available to the decision-maker, the ambiguous chance constrained problem is given by

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \bar{\mathcal{X}}_\epsilon = \Big\{\mathbf{y} \in \mathcal{X} \mid \mathbb{Q}(\mathbf{H} : f(\mathbf{y},\mathbf{H}) > 0) \leq \epsilon,\ \forall \mathbb{Q} \in \mathcal{Q}\Big\}. \end{array} \tag{3}$$

Although the problem (3) was explicitly introduced in Erdoğan and Iyengar [11], the minimax formulation has a long history in stochastic programming [28, 10, 25, 23, 24]. Motivated by the fact that the sampled problem (2) is a good approximation for the chance constrained problem (1), Erdoğan and Iyengar approximate (3) by the robust sampled problem

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \mathcal{Y}[N,\beta] = \{\mathbf{y} \in \mathcal{X} \mid f(\mathbf{y},\mathbf{z}) \leq 0, \quad \forall\ \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{H}_i^0\| \leq \beta, \quad i = 1,\ldots,N\}, \end{array} \tag{4}$$

where $\mathbf{H}_{1,N}^0$ are N IID samples drawn from the central measure $\mathbb{Q}_0$ and $\|\cdot\|$ is the norm used to define the Prohorov metric $\rho_p(\cdot,\cdot)$. For appropriately chosen norms, such problems can be solved efficiently using the technique detailed in [6]. The following approximation results were established in [11].

(a) $N \geq \frac{4d_f}{\epsilon-\mu}\log\left(\frac{12}{\epsilon-\mu}\right) + \frac{4}{\epsilon-\mu}\log\left(\frac{2}{\delta(1-\beta)}\right)$, where d_f is the VC-dimension of the function class $\{f(\mathbf{x},\cdot) : \mathbf{x} \in \mathcal{X}\}$ and $\mu = 2(\frac{\epsilon}{2} + \log(\beta + 2^{-\epsilon/2}))$, suffices to ensure $\mathbb{P}(\mathcal{Y}[N,\beta] \subseteq \mathcal{X}_\epsilon(\mathbb{Q})) \geq 1-\delta$, for any fixed $\mathbb{Q} \in \mathcal{Q}$.

(b) $N \geq \frac{2n}{\epsilon-\beta}\ln\left(\frac{2}{\epsilon-\beta}\right) + \frac{2}{\epsilon-\beta}\ln\left(\frac{1}{\delta}\right) + 2n$ ensures that the optimal solution $\widehat{\mathbf{x}}$ of (4) satisfies $\mathbb{P}(\widehat{\mathbf{x}} \in \bar{\mathcal{X}}_\epsilon) \geq 1-\delta$.

The model (1), while quite general in its ability to model constraints, is limited to the so-called single stage problems where decisions must be made before the uncertain parameter $\mathbf{H}$ is revealed. A natural extension is to consider two-stage problems where one has to commit to the first stage decision $\mathbf{x}$ *before* the realization of the uncertain parameter $\mathbf{H}$, and the second stage decision variable $\mathbf{v}$ can be chosen *after* observing the parameter $\mathbf{H}$. A prototypical example of a two-stage problem is the network design where the first stage variables are the capacities on the arcs and the second stage variables are the routing decisions. The simplest two-stage chance constrained problem is given by

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \widehat{\mathcal{X}}_\epsilon(\mathbb{Q}) = \big\{\mathbf{y} \in \mathcal{X} \mid \mathbb{Q}\big(\mathbf{H} : \exists \mathbf{v} \text{ s.t. } \mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{y},\mathbf{H})\big) > 1-\epsilon\big\}, \end{array} \tag{5}$$

where the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h}) : \mathbf{R}^n \times \mathbf{R}^m \mapsto \mathbf{R}^l$ is bi-affine, $\mathbf{v} \in \mathbf{R}^k$, and $\mathbf{W} \in \mathbf{R}^{l \times k}$. Since the matrix $\mathbf{W}$ does not depend on the realization of $\mathbf{H}$, problems of the form (5) are said to have a fixed recourse. Thus, (5) is a two-stage linear chance constrained problem with a fixed recourse. This model was introduced by Nemirovski and Shapiro [18].

One could attempt to approximate (5) by the sampled two-stage LP

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \mathcal{Y}[\mathbf{H}_{1,N}] = \{\mathbf{y} \in \mathcal{X} \mid \exists \mathbf{v}_i \text{ s.t. } \mathbf{W}\mathbf{v}_i \geq \mathbf{f}(\mathbf{y}, \mathbf{H}_i), i = 1 \ldots, N\}. \end{array} \tag{6}$$

However, note that since each scenario $\mathbf{H}_i$ has its own set of second-stage variables $\mathbf{v}_i$, $i = 1, \ldots, N$, the problem dimension grows with N, and the results of Calafiore and Campi [7, 8] no longer yield a good bound on the number of samples required to produce a good approximation. Also, since it is difficult to quantify the VC-dimension of the two-stage constraint, the results in [11] cannot be applied. Nemirovski and Shapiro [18] constructed an iterative solution algorithm for (5) that closely resembles the ellipsoid algorithm [5, 16, 26].

Our contributions in this paper are as follows.

(a) We extend the iterative algorithm proposed in [18] to solve two-stage chance constrained problems where the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ (see (5) for details) is bi-convex, i.e. convex in one variable when the other variable is held constant. We still assume a constant recourse matrix $\mathbf{W}$. This fairly straightforward extension is discussed in § 3.

(b) Since the iterative algorithm proposed in [18] closely resembles the ellipsoid algorithm, the number of iterations required to compute a feasible solution as well as the quality of the solution are functions of the radius r of a ball with the largest volume contained within the feasible set of the problem. We show that the value of the parameter r is determined by the degree of *robust* feasibility of the chance constrained problem. This is similar to the results relating the condition number of optimization problems to the complexity of solving them to optimality [20, 21, 13]. This result is proved in § 3 and discussed in § 3.1.

(c) We formulate the ambiguous two-stage chance constrained problem and modify the algorithm proposed in [18] to compute a good solution for this problem. Our extension is limited to the special case where the function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine (i.e. the case considered in [18]) and the extreme points of a certain "dual" polytope are known explicitly. We extend all the known results for chance constrained problems to the ambiguous setting.

The rest of the paper is organized as follows. In § 2 we introduce the notation that we use in the rest of the paper. In § 3 we discuss the two-stage chance constrained problems with bi-convex impact functions. In § 4 we present the extension to the ambiguous two-stage chance constrained problems. In § 5 we present the results of our numerical experiments with a two-stage network design problem on a very simple network. In § 6 we discuss the results in the paper and include some concluding remarks.

2 Notation

Sets will be denoted by calligraphic letters, e.g. $\mathcal{A}$, and $\mathcal{A}^c$ will denote the complement of the set $\mathcal{A}$. All (deterministic) vectors will be denoted by the boldface lowercase letters, e.g. $\mathbf{x}$. Random vectors and samples of random vectors will be denoted by the boldface uppercase letters, e.g. $\mathbf{H}$, and measures will be denoted by the mathematical boldface letters, e.g. $\mathbb{P}$. We will denote that a random vector $\mathbf{H}$ has distribution $\mathbb{Q}$ by $\mathbf{H} \sim \mathbb{Q}$ and a σ-algebra on a space $\mathcal{H}$ by $\mathcal{F}(\mathcal{H})$.

The norm $\|\cdot\|$ will denote the usual Euclidean norm $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{\sum_i x_i^2}$. $\mathcal{B}_r(\mathbf{y})$ will denote a Euclidean ball of radius r centered at $\mathbf{y}$, i.e. $\mathbf{B}_r(\mathbf{y}) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{y}\| \leq r\}$, and the set $\mathcal{B}_r$ will denote a Euclidean ball of radius r, i.e. $\mathcal{B}_r = \mathcal{B}_r(\mathbf{y})$ for some fixed $\mathbf{y}$. Given an input $\mathbf{x} \in \mathbf{R}^n$, a separation oracle $S_{\mathcal{A}}$ for a convex set $\mathcal{A} \subset \mathbf{R}^n$ returns an affine function $L : \mathbf{R}^n \mapsto \mathbf{R}$ satisfying

$$S_{\mathcal{A}}(\mathbf{x}) = \begin{cases} L \text{ s.t. } L(\mathbf{z}) \leq 0, \ \forall \mathbf{z} \in \mathcal{A}, \ L(\mathbf{x}) > 0 & \mathbf{x} \notin \mathcal{A}, \\ L \equiv \mathbf{0} & \text{otherwise} \end{cases}$$

The gradient of a function $L : \mathbf{R}^n \mapsto \mathbf{R}$ will be denoted by ∇L.

3 Approximating two-stage convex chance constrained problems

In this section we develop an approximation algorithm for the two-stage convex chance constrained problem

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \widehat{\mathcal{X}}_\epsilon(\mathbb{Q}) = \Big\{\mathbf{y} \in \mathcal{X} \mid \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{y}, \mathbf{H}) \notin \mathcal{C}) \leq \epsilon\Big\}, \end{aligned} \tag{7}$$

where $\mathcal{C}$ denotes the polyhedron

$$\mathcal{C} = \Big\{\mathbf{z} \in \mathbf{R}^l \mid \mathbf{z} \leq \mathbf{W}\mathbf{v}, \mathbf{v} \in \mathbf{R}^k\Big\}, \tag{8}$$

$\mathbf{x} \in \mathbf{R}^n$ is the decision vector, $\mathbf{H} \in \mathcal{H}$ is a random parameter vector distributed according to $\mathbb{Q}$, $\mathbf{W} \in \mathbf{R}^{l \times k}$ is a recourse matrix, and $\mathbf{f}(\mathbf{x}, \mathbf{h}) : \mathcal{X} \times \mathcal{H} \mapsto \mathbf{R}^l$ is an impact function. Nemirovski and Shapiro [18] introduced the particular form for the set $\mathcal{C}$ and the associated chance constrained problem (7). To reiterate, the variable $\mathbf{x}$ denotes the first stage decisions and the variable $\mathbf{v} \in \mathbf{R}^k$ denotes the second stage decisions. We assume that the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ and the sets $\mathcal{X}$ and $\mathcal{H}$ satisfy the following conditions.

Assumption 1

(i) The sets $\mathcal{X}$ and $\mathcal{H}$ are convex compact sets. Let $\|\mathbf{x}\| \leq R_{\mathcal{X}}$ (resp. $\|\mathbf{h}\| \leq R_{\mathcal{H}}$) for all $\mathbf{x} \in \mathcal{X}$ (resp. $\mathbf{h} \in \mathcal{H}$).

(ii) The impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-convex, i.e. for all fixed $\mathbf{h} \in \mathcal{H}$ (resp. $\mathbf{x} \in \mathcal{X}$), the function $\mathbf{f}(\cdot, \mathbf{h}) : \mathcal{X} \mapsto \mathbf{R}^l$ (resp. $\mathbf{f}(\mathbf{x}, \cdot) : \mathcal{H} \mapsto \mathbf{R}^l$) is a convex function.

(iii) $\mathbf{f}(\mathbf{x},\mathbf{h}) = \mathbf{f}^0(\mathbf{x}) + \mathbf{f}^1(\mathbf{x},\mathbf{h})$, *where* $\mathbf{f}^1(\mathbf{x},\alpha\mathbf{h}) = \alpha\mathbf{f}^1(\mathbf{x},\mathbf{h})$, *for all* $\alpha \geq 0$, $(\mathbf{x},\mathbf{h}) \in \mathcal{X} \times \mathcal{H}$.

(iv) There exists a non-decreasing (finite-valued) function $K_f^{\mathcal{H}} : \mathbf{R}_+ \mapsto \mathbf{R}_+$ *such that* $\|\mathbf{f}(\mathbf{x}_1,\mathbf{h}) - \mathbf{f}(\mathbf{x}_0,\mathbf{h})\| \leq K_f^{\mathcal{H}}(\|\mathbf{h}\|)\|\mathbf{x}_1 - \mathbf{x}_0\|$, *for all* $\mathbf{h} \in \mathcal{H}$ *and* $\mathbf{x}_0$, $\mathbf{x}_1 \in \mathcal{X}$. *Let* $\bar{K}_f^{\mathcal{H}} \triangleq K_f^{\mathcal{H}}(R_{\mathcal{H}})$.

(v) There exists a non-decreasing (finite-valued) function $K_f^{\mathcal{X}} : \mathbf{R}_+ \mapsto \mathbf{R}_+$ *such that* $\|\mathbf{f}(\mathbf{x},\mathbf{h}_1) - \mathbf{f}(\mathbf{x},\mathbf{h}_0)\| \leq K_f^{\mathcal{X}}(\|\mathbf{x}\|)\|\mathbf{h}_1 - \mathbf{h}_0\|$, *for all* $\mathbf{h}_1$, $\mathbf{h}_0 \in \mathcal{H}$ *and* $\mathbf{x} \in \mathcal{X}$. *Let* $\bar{K}_f^{\mathcal{X}} \triangleq K_f^{\mathcal{X}}(R_{\mathcal{X}})$.

The condition (i) is not essential for the results to hold and is almost always satisfied in practice. Assuming $\mathbf{f}(\mathbf{x},\mathbf{h})$ is convex in $\mathbf{x}$ for all fixed $\mathbf{h} \in \mathcal{H}$ is necessary to ensure that the sampled version of (7) is tractable. The assumption that $\mathbf{f}(\mathbf{x},\mathbf{h})$ is convex in $\mathbf{h}$ for all fixed $\mathbf{x} \in \mathcal{X}$ and has a component that is homogeneous allows one to leverage the concentration-of-measure property defined below in Definition 1. The assumption that $\mathbf{f}(\mathbf{x},\mathbf{h})$ is Lipschitz continuous individually in each variable is sufficient, though not necessary, to establish the approximation results. Assumption 1 may appear overly restrictive; however, there are many function classes that satisfy these. Two canonical examples are as follows.

(a) Affine constraints: $\mathbf{f}(\mathbf{x},\mathbf{h}) = \mathbf{A}_0\mathbf{x} + \sum_{i=1}^m h_i\mathbf{A}_i\mathbf{x}$ with $\mathcal{X} \subset \mathbf{R}^n$, $\mathcal{H} \subset \mathbf{R}^m$, and $\mathbf{A}_i \in \mathbf{R}^{l\times n}$ for $i = 0,\ldots,m$. The growth functions $K_f^{\mathcal{H}}(\|\mathbf{h}\|) = O(1)(|||\mathbf{A}_0||| + \|\mathbf{h}\|\sum_{i=1}^m |||\mathbf{A}_i|||)$ and $K_f^{\mathcal{X}}(\|\mathbf{x}\|) = O(1)\|\mathbf{x}\|\sum_{i=1}^m |||\mathbf{A}_i|||$, with the constants depending on the particular choice of the vector norm $\|\cdot\|$ and the matrix norm $|||\cdot|||$, satisfy Assumption 1.

(b) Second-order cone constraints: Each component of $\mathbf{f}(\mathbf{x},\mathbf{h})$ is a conic quadratic representable function [5], e.g. $f_j(\mathbf{x},\mathbf{h}) = \sqrt{(\Gamma\mathbf{A}_j\mathbf{x}+\boldsymbol{\rho})^T(\Gamma\mathbf{A}_j\mathbf{x}+\boldsymbol{\rho})} - \mathbf{t}_j^T\mathbf{x} - v_j$ with $\mathcal{X} \subset \mathbf{R}^n$, $\mathbf{A}_j \in \mathbf{R}^{k\times n}$, $\mathbf{t}_j \in \mathbf{R}^n$, $v_j \in \mathbf{R}$ for $j = 1,\ldots,l$, and $\mathcal{H} = \{\mathbf{h} = (\Gamma,\boldsymbol{\rho}) : \Gamma \in \mathbf{R}^{p\times k}, \boldsymbol{\rho} \in \mathbf{R}^p,\}$. In this case, we can set $K_f^{\mathcal{H}}(\|\mathbf{h}\|) = O(1)(|||\mathbf{T}||| + \|\mathbf{h}\||||\mathbf{A}|||)$ and $K_f^{\mathcal{X}}(\|\mathbf{x}\|) = O(1)\|\mathbf{x}\||||\mathbf{A}|||$, where $\mathbf{T} = [\mathbf{t}_1^T;\mathbf{t}_2^T;\ldots;\mathbf{t}_l^T]$, $\mathbf{A} = [\mathbf{A}_1;\mathbf{A}_2;\ldots;\mathbf{A}_l]$.

We assume that $\mathbb{Q}$ satisfies the $(\bar{\theta},\psi)$-*concentration of measure* property defined as follows [18].

Definition 1 ([18]) *Let* $\bar{\theta} \in (\frac{1}{2},1]$ *and* $\psi(\alpha,\theta) : [1,\infty) \times (\bar{\theta},1] \mapsto \mathbf{R}_+$ *be a convex, non-decreasing and non-constant function of* α. *A distribution* $\mathbb{Q}$ *on* $\mathbf{R}^m$ *is said to have* $(\bar{\theta},\psi)$*-concentration of measure property if for all* $\alpha \geq 1$ *and closed convex sets* $\mathcal{B}$ *with* $\mathbb{Q}(\mathcal{B}) \geq \theta > \bar{\theta}$,

$$\mathbb{Q}(\{\mathbf{H} \notin \alpha\mathcal{B}\}) \leq \exp\{-\psi(\alpha,\theta)\}.$$

This assumption essentially states that a small "blow-up" of the set $\mathcal{B}$ with a measure of at least $\bar{\theta}$ increases its measure exponentially. The prototypical example of a measure satisfying such a property is the multivariate Normal distribution, $\mathcal{N}(\mathbf{0},\mathbf{I})$ – it satisfies the concentration property with $\psi(\alpha,\theta) = \frac{1}{2}\alpha^2\Phi^{-1}(\theta)^2$, where $\Phi(\cdot)$ denotes the CDF of a $\mathcal{N}(0,1)$ random variable. The assumption that the impact function has a homogeneous component (see Assumption 1 part (iii)) is made to leverage the concentration property of the measure $\mathbb{Q}$.

As discussed in § 1, a convex chance constrained problem is approximately solved by computing a solution to an appropriately defined sampled problem. Erdoğan and Iyengar[11] (see also [9]) compute bounds for the number of samples required to reliably produce a solution for the chance constrained problem when the VC-dimension of the constraint function is known. Since the VC dimension of the constraint defining (7) is difficult to quantify, VC-dimension based bounds are not useful in solving (7). Calafiore and Campi [7, 8] bounded the number of samples required in terms of the number of decision variables when all the constraints are convex in $\mathbf{x}$ for any fixed $\mathbf{h}$. Since we need the second stage variables $\mathbf{v}$ to check feasibility for each sample, the number of decision variables grows linearly with the number of samples; this renders the bounds in [7, 8] worthless. We propose to approximately solve (7) using Algorithm SolveChance shown in Figure 1. SolveChance is a simple extension of an algorithm proposed by Nemirovski and Shapiro [18] to solve the special case with bi-affine impact functions $\mathbf{f}(\mathbf{x}, \mathbf{h})$. The extension to the bi-convex case is fairly straightforward; our main contribution is to show that feasibility of an appropriately defined conservative version of (7) implies that SolveChance returns a "good" solution with a high probability. Next, we carefully describe the algorithm and then prove a series of intermediate results that are needed to establish the main result.

SolveChance uses two oracles, $S_{\mathcal{X}}$ and $S_{\mathcal{R}}$. The oracle $S_{\mathcal{X}}$ is the separation oracle for the convex compact set $\mathcal{X}$ and the oracle $S_{\mathcal{R}}(\mathbf{x}; \mathbf{h})$, for a fixed $\mathbf{h} \in \mathbf{R}^m$, returns a linear inequality $L : \mathbf{R}^n \mapsto \mathbf{R}$ that separates $\mathbf{x} \in \mathbf{R}^n$ from the convex set

$$\mathcal{R} = \big\{\mathbf{x} \mid \mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}\big\} = \{\mathbf{x} \mid \exists \mathbf{v} \in \mathbf{R}^k \text{ s.t. } \mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{x}, \mathbf{h})\}. \tag{9}$$

Thus, $\mathbf{x} \in \mathcal{R}$ if, and only if, the value of the optimization problem

$$\begin{aligned} \min \quad & \|\mathbf{u} - \mathbf{x}\| \\ \text{s.t.} \quad & \mathbf{W}\mathbf{v} - \mathbf{f}(\mathbf{u}, \mathbf{h}) \geq \mathbf{0} \end{aligned} \tag{10}$$

is equal to 0. When the optimal value of (10) is strictly positive, any sub-gradient $\mathbf{d}$ at the optimal solution $\mathbf{u}^*$ satisfies $\mathbf{d}^T(\mathbf{u} - \mathbf{x}) > 0$ for all $\mathbf{u} \in \mathcal{R}$ and, therefore, serves as a separating hyperplane. Note that (10) is a convex optimization problem and can be solved very efficiently when $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is an affine function or a conic quadratic representable function [5]. The above discussion establishes that the separation oracle $S_{\mathcal{R}}(\mathbf{x}; \mathbf{h})$ is well-defined.

We will call an output $\widehat{\mathbf{x}}$ of SolveChance well-defined if $\widehat{\mathbf{x}} \neq \emptyset$. Nemirovski and Shapiro [18] established the following result (see also Theorem 4).

Theorem 1 *Suppose the output $\widehat{\mathbf{x}}$ of* SolveChance *is well-defined, (i.e. $\widehat{\mathbf{x}} \neq \emptyset$). Then $\widehat{\mathbf{x}}$ is infeasible for (7), i.e. $\widehat{\mathbf{x}} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})$, with a probability at most $M\delta$.*

Note that Theorem 1 does not consider the case $\widehat{\mathbf{x}} = \emptyset$. Next, we compute an upper bound on the probability that the output $\widehat{\mathbf{x}} = \emptyset$. Let $\mathcal{I} = \{L_1, \ldots, L_{MN}\}$ denote an ordered list of the MN linear

Input: $\epsilon > 0$, $\delta \in (0,1)$, separation oracles $S_{\mathcal{X}}$ and $S_{\mathcal{R}}$
Output: $\widehat{\mathbf{x}}$
set $\mathcal{P} \leftarrow \emptyset$, $N \leftarrow \lceil \frac{1}{\epsilon} \ln\left(\frac{1}{\delta}\right) \rceil$, $M \leftarrow \lceil 2n^2 \ln(\frac{nR_{\mathcal{X}}^2 \|\mathbf{c}\|_2}{r\omega} + 2) \rceil$
set $\mathbf{x}_0 \leftarrow \mathbf{0}$, $\mathbf{E}_0 \leftarrow R_{\mathcal{X}}\mathbf{I}$
for t = 1, ..., M **do**
    Construct a direction vector $\mathbf{d}_t$
        $s_t \leftarrow S_{\mathcal{X}}(\mathbf{x}_{t-1})$
        if $\mathbf{x}_{t-1} \notin \mathcal{X}$, **set** $\mathbf{d}_t \leftarrow \nabla s_t$
        else
            generate $\mathbf{H}_{1,N}$ IID $\mathbb{Q}$
            $L_{N(t-1)+i} \leftarrow S_{\mathcal{R}}(\mathbf{x}_{t-1}; \mathbf{H}_i)$, $i = 1, \ldots, N$.
        if $\exists\, j \in \{1, \ldots, Nt\}$ such that $L_j(\mathbf{x}_{t-1}) > 0$, **set** $\mathbf{d}_t \leftarrow \nabla L_j$
        else set $\mathbf{d}_t = \mathbf{c}$ and $\mathcal{P} \leftarrow \mathcal{P} \cup \mathbf{x}_{t-1}$
    Given $(\mathbf{x}_{t-1}, \mathbf{E}_{t-1})$ and $\mathbf{d}_t$, set $(\mathbf{x}_t, \mathbf{E}_t)$ by the Ellipsoid method update
if $\mathcal{P} = \emptyset$ **return** $\widehat{\mathbf{x}} \leftarrow \emptyset$; **else return** $\widehat{\mathbf{x}} \leftarrow \operatorname{argmin}_{\mathbf{x} \in \mathcal{P}} \{\mathbf{c}^T \mathbf{x}\}$

Figure 1: Algorithm SolveChance

inequalities generated by the calls to the oracle $S_{\mathcal{R}}$ over the course of one run of SolveChance. Let $\mathcal{X}_{\mathcal{I}}$ denote the convex compact set

$$\mathcal{X}_{\mathcal{I}} = \{\mathbf{x} \in \mathcal{X} : L_j(\mathbf{x}) \leq 0, j = 1, \ldots, MN\}. \tag{11}$$

Note that the set $\mathcal{I}$ and, therefore, $\mathcal{X}_{\mathcal{I}}$ depend on the IID samples $\mathbf{H}_{1,MN}$, where each $\mathbf{H}_i \sim \mathbb{Q}$. The following lemma was stated in [18] without a proof.

Lemma 1 *Suppose the set $\mathcal{X}_{\mathcal{I}}$ contains a Euclidean ball $\mathcal{B}_r$ of radius r. Then the output $\widehat{\mathbf{x}}$ of* SolveChance *is well defined, i.e. $\widehat{\mathbf{x}} \neq \emptyset$.*

Proof: We will prove the result by contradiction. Suppose SolveChance returns $\widehat{\mathbf{x}} = \emptyset$. Let $\{(\mathbf{x}_{t-1}, \mathbf{d}_t) : t = 1, \ldots, M\}$ denote the iterates and the separating hyperplanes generated during one run of SolveChance. Since $\widehat{\mathbf{x}} = \emptyset$, for each $\mathbf{x}_t$, $t = 0, \ldots, M-1$, either $\mathbf{x}_t \notin \mathcal{X}$ or there exists some $j \in \{1, \ldots, N(t+1)\}$ such that $L_j(\mathbf{x}_t) > 0$. Thus, $\mathbf{x}_t \notin \mathcal{X}_{\mathcal{I}}$ for all $t = 0, \ldots, M-1$.

By the definition of $\mathcal{X}_{\mathcal{I}}$, it follows that $\{(\mathbf{x}_{t-1}, \mathbf{d}_t) : t = 1, \ldots, M\}$ is a set of iterates and separating hyperplanes that could have been generated while using the Ellipsoid algorithm to solve the convex optimization problem $\min\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in \mathcal{X}_{\mathcal{I}}\}$. Since $\mathbf{x}_t \notin \mathcal{X}_{\mathcal{I}}$ for all $t = 0, \ldots, M-1$, it follows that the Ellipsoid algorithm returns an empty solution. This is a contradiction because of the choice of the iteration count M [5]. ■

For $\alpha > 1$ and $r > 0$, define

$$\mathcal{X}_\epsilon(\alpha, r) = \left\{ \mathbf{x} \in \mathcal{X}^{-r} \,\middle|\, \mathbb{Q}\Big(\mathbf{H} : \mathbf{f}(\mathbf{x}, \alpha\mathbf{H}) \in \mathcal{C}^{-r\bar{K}_f^{\mathcal{H}}}\Big) > 1 - \epsilon \right\}, \tag{12}$$

where

$$\mathcal{A}^{-\gamma} \triangleq \{\mathbf{y} \in \mathcal{A} \mid \mathbf{y} + \mathbf{u} \in \mathcal{A}, \text{ for all } \|\mathbf{u}\| \leq \gamma\}, \tag{13}$$

denote the *interior* γ-ball of the set $\mathcal{A}$. Recall that $\bar{K}_f^{\mathcal{H}} = K_f^{\mathcal{H}}(R_{\mathcal{H}})$ is the maximum value of the growth function $K_f^{\mathcal{H}}(\cdot)$. ¿From (13), it follows that

$$\mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}^{-\mu} \quad \Leftrightarrow \quad \mathbf{f}(\mathbf{x}, \mathbf{h}) + \mathbf{u} \in \mathcal{C}, \ \forall \mathbf{u} : \|\mathbf{u}\| \leq \mu. \tag{14}$$

Since $\mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}$ if, and only if, there exists a $\mathbf{v} \in \mathbf{R}^k$ such that $\mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{x}, \mathbf{h})$; we have that $\mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}^{-\mu}$ if, and only if, for all $\mathbf{u}$ with $\|\mathbf{u}\| \leq \mu$, there exists a $\mathbf{v} \in \mathbf{R}^k$ (possibly a function of $\mathbf{u}$) such that $\mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{x}, \mathbf{h}) + \mathbf{u}$. The set $\mathcal{X}_\epsilon(\alpha, r)$ can be interpreted as the set of decision vectors that are *robustly* feasible for the chance constrained set $\widehat{\mathcal{X}}_\epsilon(\mathbb{Q})$ [4].

Theorem 2 *Fix $\mathbf{y} \in \mathcal{X}_\epsilon(\alpha, r)$. Then the Euclidean ball $\mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}_\mathcal{I}$ with a probability at least $1 - MNe^{-\psi(\alpha, 1-\epsilon)}$.*

Proof: Let $\mu_r = r\bar{K}_f^{\mathcal{H}}$. Then we have that

$$\begin{aligned} 1 - \epsilon &< \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{y}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r}), \\ &= \mathbb{Q}(\alpha^{-1}\mathbf{H} : \mathbf{f}^0(\mathbf{y}) + \mathbf{f}^1(\mathbf{y}, \mathbf{H}) \in \mathcal{C}^{-\mu_r}). \end{aligned} \tag{15}$$

Let $\mathcal{H}_\mathbf{y} = \{\mathbf{h} : \mathbf{f}^0(\mathbf{y}) + \mathbf{f}^1(\mathbf{y}, \mathbf{h}) \in \mathcal{C}^{-\mu_r}\}$. Then (15) and the concentration property of $\mathbb{Q}$ imply that $\mathbb{Q}(\mathcal{H}_\mathbf{y}) \geq 1 - e^{-\psi(\alpha, 1-\epsilon)}$ provided $1 - \epsilon > \bar{\theta}$.

Since $\mathbf{y} \in \mathcal{X}^{-r}$, it follows that $\mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}$, and Lipschitz continuity of $\mathbf{f}$ implies that for all $\mathbf{x} \in \mathcal{B}_r(\mathbf{y})$ and $\mathbf{h} \in \mathcal{H}_\mathbf{y}$,

$$\|\mathbf{f}(\mathbf{x}, \mathbf{h}) - \mathbf{f}(\mathbf{y}, \mathbf{h})\| \leq K_f^{\mathcal{H}}(\|\mathbf{h}\|)\|\mathbf{x} - \mathbf{y}\| \leq rK_f^{\mathcal{H}}(R_\mathcal{H}) = \mu_r,$$

i.e. $\mathbf{f}(\mathbf{x}, \mathbf{h}) = \mathbf{f}(\mathbf{y}, \mathbf{h}) + \mathbf{u}$ for some $\mathbf{u}$ with $\|\mathbf{u}\| \leq \mu_r$. Since $\mathbf{y} \in \mathcal{X}_\epsilon(\alpha, r)$, it follows that

$$\mathbf{f}(\mathbf{x}, \mathbf{h}) \in \mathcal{C}, \ \forall \mathbf{x} \in \mathcal{B}_r(\mathbf{y}), \ \mathbf{h} \in \mathcal{H}_\mathbf{y}. \tag{16}$$

Suppose $\mathbf{H}_i \in \mathcal{H}_\mathbf{y}$, for all $i = 1, \ldots, MN$. Then (16) implies that $\mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}_\mathcal{I}$. Thus,

$$\{\mathbf{h}_{1,MN} : \mathbf{h}_i \in \mathcal{H}_\mathbf{y}, i = 1, \ldots, MN\} \subseteq \{\mathbf{h}_{1,MN} : \mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}_\mathcal{I}\}. \tag{17}$$

Consequently,

$$\begin{aligned} \mathbb{Q}^{MN}(\mathbf{H}_{1,MN} : \mathcal{B}_r(\mathbf{y}) \not\subseteq \mathcal{X}_\mathcal{I}) &\leq \mathbb{Q}^{MN}((\mathbf{H}_{1,MN} : \mathbf{H}_i \in \mathcal{H}_\mathbf{y}, i = 1, \ldots, MN)^c), \\ &\leq MN\mathbb{Q}(\mathcal{H}_\mathbf{y}^c) < MNe^{-\psi(\alpha, 1-\epsilon)}. \end{aligned}$$

■

Theorem 2 implies the following result.

Corollary 1 *Suppose $\mathcal{X}_\epsilon(\alpha, r) \neq \emptyset$. Then the output $\widehat{\mathbf{x}}$ of* SolveChance *is well-defined (i.e. $\widehat{\mathbf{x}} \neq \emptyset$) with a probability at least $1 - MNe^{-\psi(\alpha,1-\epsilon)}$.*

Proof: Lemma 1 implies that the solution $\widehat{\mathbf{x}}$ is well defined if the set $\mathcal{X}_\mathcal{I}$ contains a Euclidean ball $\mathcal{B}_r$ of radius r; and, by Theorem 2, the probability of this event is at least $1 - MNe^{-\psi(\alpha,1-\epsilon)}$. ■

Corollary 1 establishes that the output $\widehat{\mathbf{x}}$ of SolveChance is well defined with a high probability provided the chance constrained problem (7) is robustly feasible and the measure $\mathbb{Q}$ has a concentration of measure property. Next, we establish a bound on the value $\mathbf{c}^T\widehat{\mathbf{x}}$. We will call a well-defined output $\widehat{\mathbf{x}}$ (i.e. $\widehat{\mathbf{x}} \neq \emptyset$) of SolveChance an (α, r, ω)-approximation of (7) if

$$\mathbf{c}^T\widehat{\mathbf{x}} \leq \inf_{\mathbf{x}\in\mathcal{X}_\epsilon(\alpha,r)} \mathbf{c}^T\mathbf{x} + \omega. \tag{18}$$

This definition was introduced in [18].

Theorem 3 *Suppose $\mathcal{X}_\epsilon(\alpha, r) \neq \emptyset$. Then* SolveChance *returns an (α, r, ω)-approximate solution of (7) with a probability at least $(1 - MNe^{-\psi(\alpha,1-\epsilon)})(1 - M\delta)$.*

Proof: Fix $\kappa > 0$ and choose $\mathbf{y}^\kappa \in \mathcal{X}_\epsilon(\alpha, r)$ such that

$$\mathbf{c}^T\mathbf{y}^\kappa \leq \inf_{\mathbf{x}\in\mathcal{X}_\epsilon(\alpha,r)} \mathbf{c}^T\mathbf{x} + \kappa. \tag{19}$$

By Lemma 1, $\widehat{\mathbf{x}}$ is well defined on the event $\{\mathcal{B}_r(\mathbf{y}^\kappa) \subseteq \mathcal{X}_\mathcal{I}\}$ and by Theorem 2 the probability of this event is at least $1 - MNe^{-\psi(\alpha,1-\epsilon)}$. Combining this with Theorem 1, we can show that SolveChance produces a well-defined output $\widehat{\mathbf{x}}$ that is feasible for (7) with a probability at least $(1 - MNe^{-\psi(\alpha,1-\epsilon)})(1 - M\delta)$.

Next, we show that $\mathcal{B}_r(\mathbf{y}^\kappa) \subseteq \mathcal{X}_\mathcal{I}$ implies

$$\mathbf{c}^T\widehat{\mathbf{x}} \leq \min_{\mathbf{x}\in\mathcal{X}_\mathcal{I}} \mathbf{c}^T\mathbf{x} + \omega. \tag{20}$$

This result is established by exploiting the close resemblance of SolveChance to the Ellipsoid algorithm. We closely follow the analysis of the Ellipsoid algorithm detailed in the proof of Theorem 5.2.1 in [5].

Let $\mathbf{x}^*_\mathcal{I} = \text{argmin}_{\mathbf{x}\in\mathcal{X}_\mathcal{I}}\{\mathbf{c}^T\mathbf{x}\}$. Let $(\mathbf{x}_t, \mathbf{E}_t)$, $t = 0, 1, \ldots, M-1$, denote the iterates generated by SolveChance . Let $\mathcal{E}_t$ denote the ellipsoid $\mathcal{E}_t = \{\mathbf{z} \mid (\mathbf{z}-\mathbf{x}_t)^T\mathbf{E}_t^{-1}(\mathbf{z}-\mathbf{x}_t) \leq 1\}$. The choice of the iteration count M ensures that there exists $\nu \leq 1$ and $\mathbf{z} \in \mathcal{X}_\mathcal{I}$ such that

(i) $\mathbf{vol}(\mathcal{E}_t)/r^n \leq \nu \leq 1$,

(ii) $\mathbf{y} = \mathbf{x}^*_\mathcal{I} + \nu(\mathbf{z} - \mathbf{x}^*_\mathcal{I}) \notin \mathcal{E}_M$.

Since $\mathbf{y} \in \mathcal{X}_\mathcal{I} \subset \mathcal{X}$, it follows that $\mathbf{y} \in \mathcal{E}_0$. Therefore, there exists $\tau < M$ such that $\mathbf{y} \in \mathcal{E}_\tau$ but $\mathbf{y} \notin \mathcal{E}_{\tau+1}$. Then it follows that $\mathbf{d}_\tau^T\mathbf{s} > \mathbf{d}_\tau^T\mathbf{x}_\tau$ for all $\mathbf{s} \in \mathcal{E}^c_{\tau+1} \cap \mathcal{E}_\tau$, and, in particular,

$$\mathbf{d}_\tau^T\mathbf{y} > \mathbf{d}_\tau^T\mathbf{x}_\tau. \tag{21}$$

We claim that the iterate $\mathbf{x}_\tau \in \mathcal{P}$, i.e. it is one of the candidate points for computing the output $\widehat{\mathbf{x}}$. Suppose this is not the case. Then we must have that $\mathbf{x}_\tau \notin \mathcal{X}_\tau$, where

$$\mathcal{X}_\tau = \{\mathbf{x} \in \mathcal{X} \mid L_j(\mathbf{x}) \leq 0, j = 1, \ldots, N(\tau+1)\} \supseteq \mathcal{X}_\mathcal{I},$$

and the separating hyperplane $\mathbf{d}_\tau$ must satisfy $\mathbf{d}_\tau^T \mathbf{s} \leq \mathbf{d}_\tau^T \mathbf{x}_\tau$ for all $\mathbf{s} \in \mathcal{X}_\tau$. Thus, the bound (21) together with the fact that $\mathbf{y} \in \mathcal{X}_\mathcal{I} \subseteq \mathcal{X}_\tau$ leads to a contradiction. Now, the analysis in the proof of Theorem 5.2.1 in [5] implies that $\mathbf{c}^T \mathbf{x}_\tau \leq \min_{\mathbf{x} \in \mathcal{X}_\mathcal{I}} \mathbf{c}^T \mathbf{x} + \omega$ and the bound (20) follows.

Finally, $\mathcal{B}_r(\mathbf{y}^\kappa) \subseteq \mathcal{X}_\mathcal{I}$ implies $\min_{\mathbf{x} \in \mathcal{X}_\mathcal{I}} \mathbf{c}^T \mathbf{x} \leq \mathbf{c}^T \mathbf{y}^\kappa$, which together with (19) imply that

$$\mathbf{c}^T \widehat{\mathbf{x}} \leq \min_{\mathbf{x} \in \mathcal{X}_\mathcal{I}} \mathbf{c}^T \mathbf{x} + \omega \leq \mathbf{c}^T \mathbf{y}^\kappa + \omega \leq \inf_{\mathbf{x} \in \mathcal{X}_\epsilon(\alpha, r)} \mathbf{c}^T \mathbf{x} + \kappa + \omega$$

Since $\kappa > 0$ was arbitrary, the result follows. ■

3.1 Discussion of the approximation result

Algorithm SolveChance has three tunable parameters, namely δ, ω and r. Nemirovski and Shapiro [18] study the effects of these parameters on the running time and approximation quality of SolveChance. While the parameters δ and ω have a well-defined meaning, the parameter r is rather ad-hoc and it is not clear how to set its value. The parameter r is clearly very important for the performance of SolveChance: the iteration count M, the probability that the output $\widehat{\mathbf{x}}$ of SolveChance is well-defined and feasible for (7) (see Theorem 1 and Corollary 1), and the approximation guarantee on the output $\widehat{\mathbf{x}}$ (see Theorem 3) are all inversely proportional to the parameter r. Since the set $\mathcal{X}_\mathcal{I}$ is random, selecting r is difficult. Yet, SolveChance requires r as an input.

The main contribution of this paper is to provide guidance in selecting r. The iteration count M as well as the probability that a well-defined $\widehat{\mathbf{x}}$ is infeasible are inversely proportional to r. This suggests that r should be set as high as possible. For any fixed $\alpha > 0$, the maximum allowed value $\bar{r}$ of r is limited by the requirement that the set $\mathcal{X}_\epsilon(\alpha, r)$ (see (12)) is non-empty, i.e. the set of decisions that are *robustly* feasible for (7) is non-empty [4]. Thus, the computational complexity of SolveChance is intimately related to the robust feasibility of (7) – the more robust the chance constrained problem, the easier it is to compute a feasible solution. This is similar to the relationship between the complexity of computing an optimal solution of a conic linear program and its condition number [20, 21, 13]. Although setting a high value for the parameter r induces SolveChance to efficiently produce a feasible solution, it results in a weak approximation guarantee (see (18)).

Let $p = (1 - M\delta)(1 - MNe^{-\psi(\alpha, 1-\epsilon)})$ denote the probability that the output $\widehat{\mathbf{x}}$ is well-defined and feasible. If $p > 0$, then $T = \frac{-\ln(\gamma)}{p}$ independent replications of SolveChance ensure that with a probability $1 - \gamma$ at least one of the outputs is feasible. The requirement $p > 0$ implies an upper bound on M, and consequently, a lower bound $\underline{r}$ on r. Thus, it follows that there is a lower bound on the achievable approximation guarantee. Within the range $[\underline{r}, \bar{r}]$ one can trade-off the optimality with efficiency.

4 Ambiguous two-stage chance constrained problems

In this section, we extend the approximation results to ambiguous two-stage chance constrained problems where the distribution of the uncertain parameter $\mathbf{H}$ is not completely known; instead, the limited knowledge about the distribution is characterized by the uncertainty set

$$\mathcal{Q} = \big\{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\big\}. \tag{22}$$

The metric $\rho_p(\cdot,\cdot)$ denotes the Prohorov metric defined as

$$\rho_p(\mathbb{Q}_1, \mathbb{Q}_2) = \inf\left\{\gamma : \mathbb{Q}_1(\mathcal{B}) \leq \mathbb{Q}_2(\mathcal{B}^\gamma) + \gamma, \forall \mathcal{B} \in \mathcal{F}(\mathcal{H})\right\}, \tag{23}$$

where

$$\mathcal{B}^\gamma = \Big\{\mathbf{x} \in \mathcal{X} : \inf_{\mathbf{z}\in\mathcal{B}} \|\mathbf{x} - \mathbf{z}\| \leq \gamma\Big\}.$$

Although the definition appears asymmetric, ρ_p is a metric. It plays an important role in probability because it metrizes weak convergence. Moreover, $\rho_p(\mathbb{Q}_1, \mathbb{Q}_2)$ is the minimum distance "in probability" between random variables distributed according to $\mathbb{Q}_i$, $i = 1, 2$.

The assumption here is that the uncertain parameter $\mathbf{H}$ is distributed according to some *fixed* distribution $\mathbb{Q} \in \mathcal{Q}$; however, the decision maker can only estimate the distribution to within the error β. The goal is to compute a solution $\widehat{\mathbf{x}}$ that performs "well" for all distributions in the set $\mathcal{Q}$. We will characterize the details of the approximation later in this section. We make the following additional assumptions on the function $\mathbf{f}(\mathbf{x}, \mathbf{h})$.

Assumption 2

(a) The function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine, i.e. $\mathbf{f}(\mathbf{x}, \mathbf{h}) = \mathbf{a}_0(\mathbf{x}) + \mathbf{A}_1(\mathbf{x})\mathbf{h}$, where the vector $\mathbf{a}_0(\mathbf{x})$ and the matrix $\mathbf{A}_1(\mathbf{x})$ are affine functions of $\mathbf{x}$.

(b) The extreme points $\{\bar{\boldsymbol{\lambda}}^{(i)} : i = 1, \ldots, p\}$ of the polytope $\{\boldsymbol{\lambda} : \mathbf{W}^T\boldsymbol{\lambda} = \mathbf{0}, \mathbf{1}^T\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq 0\}$ are explicitly known.

These constraints are quite restrictive and we will comment on them in § 5.

The aim of this section is to establish that Algorithm SolveAmbChance displayed in Figure 2 produces a solution $\widehat{\mathbf{x}}$ that performs "well" for all $\mathbb{Q} \in \mathcal{Q}$. The structure of SolveAmbChance is very similar to SolveChance, with the following two distinctions. First, the number of samples N per iteration is given by $N = \lceil \frac{1}{\epsilon-\beta} \ln\left(\frac{1}{\delta}\right)\rceil$ (contrast with $N = \lceil \frac{1}{\epsilon} \ln\left(\frac{1}{\delta}\right)\rceil$); thus ambiguity requires us to generate more samples per iteration. And second, instead of $S_{\mathcal{R}}$, Algorithm SolveAmbChance employs the oracle $S_{\mathcal{R}_\beta}$ that is a separation oracle for the set (for a fixed $\mathbf{h}$)

$$\begin{aligned} \mathcal{R}_\beta &= \Big\{\mathbf{x} \mid \mathbf{f}(\mathbf{x}, \mathbf{z}) \in \mathcal{C},\ \forall\ \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{h}\| \leq \beta\Big\}, \\ &= \Big\{\mathbf{x} \mid\ \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{h}\| \leq \beta,\ \exists \mathbf{v} \text{ s.t. } \mathbf{W}\mathbf{v} \geq \mathbf{f}(\mathbf{x}, \mathbf{z})\Big\}. \end{aligned} \tag{24}$$

Input: $\epsilon > 0$, $\delta \in (0,1)$, separation oracles $S_{\mathcal{X}}$ and $S_{\mathcal{R}_\beta}$
Output: $\widehat{\mathbf{x}}$
set $\mathcal{P} \leftarrow \emptyset$, $N = \lceil \frac{1}{\epsilon-\beta} \ln\left(\frac{1}{\delta}\right) \rceil$, $M = \lceil 2n^2 \ln(\frac{nR_{\mathcal{X}}^2 \|\mathbf{c}\|_2}{r\omega} + 2) \rceil$
set $\mathbf{x}_0 \leftarrow \mathbf{0}$, $\mathbf{E}_0 \leftarrow R_{\mathcal{X}}\mathbf{I}$
for t = 1, ..., M **do**
  Construct a direction vector $\mathbf{d}_t$
    $s_t \leftarrow S_{\mathcal{X}}(\mathbf{x}_{t-1})$
    if $\mathbf{x}_{t-1} \notin \mathcal{X}$, **set** $\mathbf{d}_t \leftarrow \nabla s_t$
    else
      generate $\mathbf{H}_{1,N}$ IID $\mathbb{Q}_0$
      $L_{N(t-1)+i} \leftarrow S_{\mathcal{R}_\beta}(\mathbf{x}_{t-1}; \mathbf{H}_i)$, $i = 1, \ldots, N$.
    if $\exists\, j \in \{1, \ldots, Nt\}$ such that $L_j(\mathbf{x}_{t-1}) > 0$, **set** $\mathbf{d}_t \leftarrow \nabla L_j$
    else set $\mathbf{d}_t = \mathbf{c}$ and $\mathcal{P} \leftarrow \mathcal{P} \cup \mathbf{x}_{t-1}$
  Given $(\mathbf{x}_{t-1}, \mathbf{E}_{t-1})$ and $\mathbf{d}_t$, set $(\mathbf{x}_t, \mathbf{E}_t)$ by the Ellipsoid method update
if $\mathcal{P} = \emptyset$ **return** $\widehat{\mathbf{x}} \leftarrow \emptyset$; **else return** $\widehat{\mathbf{x}} \leftarrow \text{argmin}_{\mathbf{x} \in \mathcal{P}} \{\mathbf{c}^T \mathbf{x}\}$

Figure 2: Algorithm SolveAmbChance

The set $\mathcal{R}_\beta$ has the same structure as the feasible set of an adjustably robust linear program [3]. It is well-known that checking membership in such a set is NP-Complete when the underlying polytope is described by a set of inequalities [3]. Since we assume (see Assumption 2-(b)) that the extreme points of the underlying polytope are explicitly available, membership in $\mathcal{R}_\beta$ can be checked efficiently. From (24), it follows $\mathbf{x} \in \mathcal{R}_\beta$ if, and only if, for all $\mathbf{z}$ satisfying $\|\mathbf{z} - \mathbf{h}\| \leq \beta$,

$$\begin{array}{rcrcl} 0 & \leq & P_{xz} & = & \max \quad \theta \\ & & & & \text{s.t.} \quad \mathbf{W}\mathbf{v} - \mathbf{f}(\mathbf{x}, \mathbf{z}) \geq \mathbf{1}\theta. \end{array} \tag{25}$$

It is easy to check that P_{xz} is always feasible. In order to construct the separating hyperplane, we consider the following two cases.

(i) There exists $\mathbf{y}$ such that $\mathbf{W}\mathbf{y} > \mathbf{0}$. In this case, $\mathbf{x} \in \mathcal{R}_\beta$ for all $\mathbf{h}$. Thus, $S_{\mathcal{R}_\beta}(\mathbf{x}) = \mathbf{0}$.

(ii) There does not exist $\mathbf{y}$ such that $\mathbf{W}\mathbf{y} > \mathbf{0}$. Then P_{xz} is bounded; thus, by strong duality, there is no duality gap, i.e.

$$\begin{aligned} P_{xz} &= \min\{-\big(\mathbf{f}(\mathbf{x}, \mathbf{z})\big)^T \boldsymbol{\lambda} : \mathbf{W}^T \boldsymbol{\lambda} = \mathbf{0}, \mathbf{1}^T \boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \mathbf{0}\}, \\ &= \min\{-\big(\mathbf{f}(\mathbf{x}, \mathbf{z})\big)^T \bar{\boldsymbol{\lambda}}^{(i)} : i = 1, \ldots, p\}, \end{aligned}$$

where $\{\bar{\boldsymbol{\lambda}}^{(i)} : i = 1, \ldots, p\}$ are the extreme points of the dual polytope. Recall that we have assumed that $\{\bar{\boldsymbol{\lambda}}^{(i)}\}$ are explicitly known.

Thus, $\mathbf{x} \in \mathcal{R}_\beta$ if, and only if,

$$\begin{aligned} 0 &\leq \min\{P_{xz} : \|\mathbf{z}-\mathbf{h}\| \leq \beta\}, \\ &= \min_{1\leq i\leq p} \left\{ -(\mathbf{a}_0(\mathbf{x}) + \mathbf{A}_1(\mathbf{x})\mathbf{h})^T \bar{\boldsymbol{\lambda}}^{(i)} - \beta \|\mathbf{A}_1(\mathbf{x})^T \bar{\boldsymbol{\lambda}}^{(i)}\| \right\}. \end{aligned}$$

Suppose the minimum above is strictly negative. Let $k = \text{argmin}_{1\leq i\leq p} \left\{ -(\mathbf{a}_0(\mathbf{x}) + \mathbf{A}_1(\mathbf{x})\mathbf{h})^T \bar{\boldsymbol{\lambda}}^{(i)} - \beta \|\mathbf{A}_1(\mathbf{x})^T \bar{\boldsymbol{\lambda}}^{(i)}\| \right\}$. For a fixed $\mathbf{h}$ consider the convex function $g(\mathbf{u};\mathbf{h}) : \mathbf{R}^n \mapsto \mathbf{R}$ defined as follows

$$g(\mathbf{u};\mathbf{h}) = (\mathbf{a}_0(\mathbf{u}) + \mathbf{A}_1(\mathbf{u})\mathbf{h})^T \bar{\boldsymbol{\lambda}}^{(k)} + \beta \|\mathbf{A}_1(\mathbf{u})^T \bar{\boldsymbol{\lambda}}^{(k)}\|.$$

Then any sub-gradient of the function $g(\cdot;\mathbf{h})$ at $\mathbf{u} = \mathbf{x}$ serves as the separating hyperplane.

The following result extends Theorem 1 to the ambiguous setting. Note that for all the results in this section the relevant probability measure is the product measure $\mathbb{Q}_0^{MN}$, since all the samples are drawn independently from the central measure $\mathbb{Q}_0$.

Theorem 4 *Suppose the output $\widehat{\mathbf{x}}$ of* SolveAmbChance *is well-defined, i.e. $\widehat{\mathbf{x}} \neq \emptyset$. Then, for every fixed $\mathbb{Q} \in \mathcal{Q}$, we have that $\mathbb{Q}(\mathbf{H} : \mathbf{f}(\widehat{\mathbf{x}}, \mathbf{H}) \notin \mathcal{C}) > \epsilon$ with a probability at most $M\delta$.*

Proof: Fix a measure $\mathbb{Q} \in \mathcal{Q}$ and let $\widehat{\mathcal{X}}_\epsilon(\mathbb{Q}) = \{\mathbf{x} \in \mathcal{X} \mid \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{x}, \mathbf{H}) \notin \mathcal{C}) \leq \epsilon\}$. By construction, the events $\{\widehat{\mathbf{x}} \neq \emptyset\} = \cup_{t=0}^{M-1} \{\mathbf{x}_t \in \mathcal{P}\}$. Consequently,

$$\{\widehat{\mathbf{x}} \neq \emptyset\} \cap \{\widehat{\mathbf{x}} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\} \subseteq \cup_{t=0}^{M-1} \big(\{\mathbf{x}_t \in \mathcal{P}\} \cap \{\mathbf{x}_t \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\} \big). \tag{26}$$

Fix t. Let B_t denote the event that $\mathbf{x}_{t-1}$ satisfies all the N inequalities generated by the oracle S_{R_β} at iteration t. Define

$$\mathcal{Y}_t[N,\beta] = \left\{ \mathbf{x} \in \mathcal{X} \mid \mathbf{f}(\mathbf{x},\mathbf{z}) \in \mathcal{C},\ \forall\ \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{H}_i^0\| \leq \beta, i = N(t-1)+1, \ldots, Nt \right\}, \tag{27}$$

where $\mathbf{H}^0_{N(t-1)+1,Nt}$ denote N IID samples drawn according to the central probability measure $\mathbb{Q}_0$ at iteration t. Then it is clear that the event $B_t = \{\mathbf{x}_{t-1} \in \mathcal{Y}_t[N,\beta]\}$.

Let A_t denote the event that the iterate $\mathbf{x}_{t-1}$ satisfies all the $N(t-1)$ inequalities generated by the oracle $S_{\mathcal{R}_\beta}$ *before* iteration t. Then it is clear that $\{\mathbf{x}_{t-1} \in \mathcal{P}\} = A_t \cap B_t$. Thus,

$$\begin{aligned} \mathbb{Q}_0^{Nt}\big(A_t \cap B_t \cap \{\mathbf{x}_{t-1} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\}\big) &\leq \mathbb{Q}_0^{Nt}\big(B_t \cap \{\mathbf{x}_{t-1} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\}\big), \\ &\leq \mathbb{Q}_0^{Nt}\big(B_t \mid \{\mathbf{x}_{t-1} \notin \widehat{\mathcal{X}}_\epsilon(\mathbb{Q})\big), \\ &\leq \delta, \end{aligned} \tag{28}$$

where the bound (28) follows from Theorem 6 in [11]. The result follows applying the union bound to the expression in (26). ■

As before, define

$$\mathcal{X}_\mathcal{I} = \{\mathbf{x} \in \mathcal{X} : L_j(\mathbf{x}) \leq 0, j = 1, \ldots, MN\},$$

where L_j, $j = 1, \ldots, MN$, denote the set of linear inequalities generated by the oracle $S_{\mathcal{R}_\beta}$ over the course of Algorithm SolveAmbChance. Then a simple extension of the technique used to prove Lemma 1 establishes the following.

Lemma 2 *Suppose the set $\mathcal{X}_\mathcal{I}$ contains a Euclidean ball $\mathcal{B}_r$ of radius r. Then the solution $\widehat{\mathbf{x}}$ returned by the* SolveAmbChance *is well defined, i.e. $\widehat{\mathbf{x}} \neq \emptyset$.*

For $\alpha > 1$, $r > 0$, and a fixed measure $\mathbb{Q} \in \mathcal{Q}$, let

$$\mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r) = \left\{ \mathbf{x} \in \mathcal{X}^{-r} \,\middle|\, \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{x}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r}) > 1 - \epsilon \right\}, \tag{29}$$

where

$$\mu_r = (\alpha + 1)\beta K_f^{\mathcal{X}}\big(R_\mathcal{X}\big) + r K_f^{\mathcal{H}}\big(R_\mathcal{H} + \beta\big). \tag{30}$$

The set $\mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r)$ denotes the set of points that are robustly feasible for the chance constraint corresponding to the measure $\mathbb{Q}$.

Theorem 5 *Fix $\mathbf{y} \in \mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r)$. Then the Euclidean ball $\mathcal{B}_r(\mathbf{y}) \subseteq \mathcal{X}_\mathcal{I}$ with a probability at least $1 - MNe^{-\psi(\alpha, 1-\epsilon-\beta)}$.*

Proof: Let $\mu_r = (\alpha + 1)\beta K_f^{\mathcal{X}}\big(R_\mathcal{X}\big) + r K_f^{\mathcal{H}}\big(R_\mathcal{H} + \beta\big)$. Then, we have that

$$\begin{aligned}
1 - \epsilon &< \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{y}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r}), && (31)\\
&\leq \beta + \mathbb{Q}_0(\mathbf{H} + \mathbf{u} : \|\mathbf{u}\| \leq \beta, \mathbf{f}(\mathbf{y}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r}), && (32)\\
&\leq \beta + \mathbb{Q}_0(\mathbf{H} : \mathbf{f}(\mathbf{y}, \alpha\mathbf{H}) \in \mathcal{C}^{-\mu_r + \alpha\beta K_f^{\mathcal{X}}(R_\mathcal{X})}), && (33)\\
&= \beta + \mathbb{Q}_0(\alpha^{-1}\mathbf{H} : \mathbf{f}^0(\mathbf{y}) + \mathbf{f}^1(\mathbf{y}, \mathbf{H}) \in \mathcal{C}^{-\mu_r + \alpha\beta K_f^{\mathcal{X}}(R_\mathcal{X})}), && (34)
\end{aligned}$$

where the inequalities (31) and (32), respectively, follow from the definitions of $\mathbf{y}$ and the Prohorov metric, and the inequality (33) follows from the Lipschitz continuity of the function f.

Let $\mathcal{H}_\mathbf{y} = \{\mathbf{h} : \mathbf{f}^0(\mathbf{y}) + \mathbf{f}^1(\mathbf{y}, \mathbf{h}) \in \mathcal{C}^{-\mu_r + \alpha\beta K_f^{\mathcal{X}}(R_\mathcal{X})}\}$. Then (34) and the concentration property of $\mathbb{Q}_0$ imply that $\mathbb{Q}_0(\mathcal{H}_\mathbf{y}) \geq 1 - e^{-\psi(\alpha, 1-\epsilon-\beta)}$ provided $1 - \epsilon - \beta > \bar{\theta}$. By Assumption 1 we have

$$\mathcal{H}_\mathbf{y} \subseteq \left\{ \mathbf{h} : \mathbf{f}(\mathbf{y}, \mathbf{z}) \in \mathcal{C}^{-\mu_r + (\alpha+1)\beta K_f^{\mathcal{X}}(R_\mathcal{X})}, \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{h}\| \leq \beta \right\}. \tag{35}$$

Thus, for all $\mathbf{h} \in \mathcal{H}_\mathbf{y}$ and $\mathbf{x} \in \mathcal{B}_r(\mathbf{y})$, Assumption 1 implies that for all $\mathbf{z}$ satisfying $\|\mathbf{z} - \mathbf{h}\| \leq \beta$, we have that

$$\begin{aligned}
\|\mathbf{f}(\mathbf{x}, \mathbf{z}) - \mathbf{f}(\mathbf{y}, \mathbf{z})\| &\leq K_f^{\mathcal{H}}(\|\mathbf{z}\|)\|\mathbf{x} - \mathbf{y}\|,\\
&\leq r K_f^{\mathcal{H}}(\|\mathbf{h}\| + \beta),\\
&\leq r K_f^{\mathcal{H}}(R_\mathcal{H} + \beta). && (36)
\end{aligned}$$

Since $\mathbf{h} \in \mathcal{H}_\mathbf{y}$, (35) and (36) imply that for all $\mathbf{z}$ satisfying $\|\mathbf{z} - \mathbf{h}\| \leq \beta$,

$$\mathbf{f}(\mathbf{x}, \mathbf{z}) \in \mathcal{C}^{-\mu_r + (\alpha+1)\beta K_f^{\mathcal{X}}(R_\mathcal{X}) + r K_f^{\mathcal{H}}(R_\mathcal{H} + \beta)} = \mathcal{C}^0 = \mathcal{C}.$$

Consequently,

$$\begin{aligned}
&\mathbb{Q}_0^{MN}(\mathbf{H}_{1,MN} : \mathcal{B}_r(\mathbf{y}) \not\subseteq \mathcal{X}_{\mathcal{I}}) \\
&\quad \leq \quad \mathbb{Q}_0^{MN}((\mathbf{H}_{1,MN} : \mathbf{f}(\mathbf{x},\mathbf{z}) \in \mathcal{C}, \forall \mathbf{x} \in \mathcal{B}_r(\mathbf{y}), \forall \mathbf{z} \text{ s.t. } \|\mathbf{z} - \mathbf{H}_i\| \leq \beta, i = 1, \ldots, MN)^c), \\
&\quad \leq \quad MN\mathbb{Q}_0(\mathcal{H}_{\mathbf{y}}^c) \leq MNe^{-\psi(\alpha, 1-\epsilon-\beta)}.
\end{aligned}$$

■

The following corollary establishes that the output of SolveAmbChance is well-defined with a high probability.

Corollary 2 *Suppose $\mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r) \neq \emptyset$. Then the output $\widehat{\mathbf{x}}$ of* SolveAmbChance *is well-defined (i.e. $\widehat{\mathbf{x}} \neq \emptyset$) with a probability at least $1 - MNe^{-\psi(\alpha,1-\epsilon-\beta)}$.*

Proof: Lemma 2 implies that the solution $\widehat{\mathbf{x}}$ is well defined if the set $\mathcal{X}_{\mathcal{I}}$ contains a Euclidean ball $\mathcal{B}_r$ of radius r and the probability of such event is, by Theorem 5, at least $1 - MNe^{-\psi(\alpha,1-\epsilon-\beta)}$. ■

Remark 1 *By setting $\beta = 0$, we recover the corresponding "unambiguous" versions of Theorem 4 and Corollary 2, namely Theorem 1 and Corollary 1.*

We are now in position to state the main result of this section. Consider the chance constrained problem,

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x} \in \widehat{\mathcal{X}}_\epsilon(\mathbb{Q}) = \Big\{\mathbf{y} \in \mathcal{X} \mid \mathbb{Q}(\mathbf{H} : \mathbf{f}(\mathbf{y}, \mathbf{H}) \notin \mathcal{C}) \leq \epsilon\Big\}, \end{array} \tag{37}$$

corresponding to a measure $\mathbb{Q} \in \mathcal{Q}$. We will call $\widehat{\mathbf{x}}$ an $(\mathbb{Q}, \alpha, r, \omega)$-approximate solution of (37) if

$$\mathbf{c}^T\widehat{\mathbf{x}} \leq \inf_{\mathbf{x} \in \mathcal{X}_\epsilon(\mathbb{Q},\alpha,r)} \mathbf{c}^T\mathbf{x} + \omega. \tag{38}$$

Then the following result holds.

Theorem 6 *For all $\mathbb{Q} \in \mathcal{Q}$ such that $\mathcal{X}_\epsilon(\mathbb{Q}, \alpha, r) \neq \emptyset$,* SolveAmbChance *returns an $(\mathbb{Q}, \alpha, r, \omega)$-approximate solution with a probability at least $(1 - MNe^{-\psi(\alpha,1-\epsilon-\beta)})(1 - M\delta)$.*

The proof of this result proceeds along the lines of Theorem 3.

5 Computational Experiments

In this section we illustrate the techniques proposed in this paper by solving a two-stage network design on the simple network displayed in Figure 3. The node s is a source node with an infinite capacity and the nodes 1,2, and 3 are sink nodes with demands given by the vector $\mathbf{d} = (d_1, d_2, d_3)^T \geq \mathbf{0}$. For a given demand vector $\mathbf{d}$ the network design problem is given by

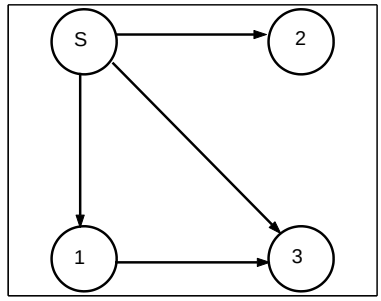

Figure 3: Chance constrained Network Design

$$\begin{array}{lrcl} \min & \mathbf{c}^T\mathbf{u} & & \\ \text{s.t.} & \mathbf{A}\mathbf{v} & \geq & \mathbf{d}, \\ & \mathbf{u}-\mathbf{v} & \geq & \mathbf{0}, \\ & \mathbf{u},\mathbf{v} \geq \mathbf{0}, & & \end{array} \tag{39}$$

where the cost vector $\mathbf{c}$ is assumed to be strictly positive, $-\mathbf{A}$ denotes the node-arc incidence matrix of the network, the variable $\mathbf{u}$ denotes the capacity on the arcs, and the variable $\mathbf{v}$ denotes the flow on the network.

Once the network is constructed, i.e. the capacities $\mathbf{u}$ are installed on the arc, we assume that it will be used over a reasonably long period over which the demand $\mathbf{d}$ can change. We model variation in demand by assuming that it is random. In particular we assume that the random demand $\mathbf{D} \sim \mathbb{Q}$ where $\mathbb{Q} \in \mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}_0, \mathbb{Q}) \leq \beta\}$, $\mathbb{Q}_0 = \mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$ is a multivariate normal distribution with mean $\bar{\mathbf{d}}$ and covariance matrix $\sigma^2\mathbf{I}$, and $\mathbf{I}$ denotes the Identity matrix. Note that the flow conservation constraints $\mathbf{A}\mathbf{v} \geq \mathbf{d}$ in (39) are formulated as inequalities instead of equalities as is usually the case – this is necessary to accommodate random demands $\mathbf{D}$.

Although the capacity cannot, typically, be altered over the life of the network, the flow $\mathbf{v}$ is chosen *after* observing the realization of the demand. Thus, the network design problem is a two-stage optimization problem: the capacities $\mathbf{u}$ are the first stage variables, and the flows $\mathbf{v} \in \mathbf{R}^n$ are the second stage variables. The objective of the two-stage optimization problem is to find a minimal cost capacity allocation $\mathbf{u}$ that guarantees that at least $(1-\epsilon)$-fraction of the random demand $\mathbf{D}$ can be feasibly routed in the resulting network. One can transform this network design problem into a chance constrained problem of the form (7) by setting

$$\mathbf{x} = \mathbf{u}, \quad \mathbf{H} = \mathbf{D}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{A} \\ -\mathbf{I} \\ \mathbf{I} \end{bmatrix}, \quad f(\mathbf{x}, \mathbf{H}) = \begin{bmatrix} \mathbf{H} \\ -\mathbf{x} \\ \mathbf{0} \end{bmatrix}, \quad \mathcal{X} = \mathbf{R}^n_+.$$

In our computational experiments, we compare performances of three different solution strategies:

(a) Deterministic solution: $\mathbf{x}_{det}$ denotes a solution of the deterministic optimization problem (39) with the demand vector $\mathbf{D}$ is set equal to its mean value $\bar{\mathbf{d}}$.

(b) Chance constrained solution: $\mathbf{x}_{sc}$ is an output of Algorithm SolveChance .

(c) Ambiguous chance constrained solution: $\mathbf{x}_{sac}$ is an output of Algorithm SolveAmbChance .

Instead of simply verifying the theoretical results presented in the previous sections, we have attempted to investigate issues that we were not able to settle theoretically. For example, we test the hypothesis that the ambiguous chance constrained solution is very conservative (and protects against a set of measures much larger than $\mathcal{Q}$) by using different test distributions $\mathbb{Q}$.

5.1 Algorithmic details

Oracles and Sampling SolveChance uses two oracles, $S_{\mathcal{X}}$ and $S_{\mathcal{R}}$, and SolveAmbChance exploits $S_{\mathcal{X}}$ and $S_{\mathcal{R}_\beta}$. Since $\mathcal{X} = \mathbf{R}^n_+$ and the function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine, (10) implies that $S_{\mathcal{X}}$ and $S_{\mathcal{R}}$ are polynomial-time oracles. The oracle $S_{\mathcal{R}_\beta}$ requires explicit characterization of extreme points of a polytope, therefore is not polynomial for some cases. However, for a network design problem, Atamtürk and Zhang [2] establish that $S_{\mathcal{R}_\beta}$ is a polynomial-time oracle for some special networks.

The algorithms SolveChance and SolveAmbChance use samples from $\mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$. Since SolveAmbChance uses more samples than SolveChance , to avoid generating too many samples, we first run SolveAmbChance and use the samples generated during its run to calculate an output of SolveChance .

Stopping criterion The number of iterations M in SolveChance and SolveAmbChance (see Figure 1 or Figure 2) was chosen large enough to guarantee convergence in the underlying Ellipsoid-like algorithm. In fact, any stopping rule that guarantees convergence in the Ellipsoid algorithm is sufficient for the results in this paper to hold.

In our experiments we use a new stopping rule proposed in [5]. Let M denote the upper bound defined in Figure 1. For $t = 1, \ldots, M$, define $v_t = R_{\mathcal{X}}\|\mathbf{d}_t\| - \mathbf{d}_t^T\mathbf{x}_{t-1}$, $V_t = \max\{V_{t-1}, v_t\}$, with $V_0 = 0$, and $\rho_t = |\,\mathbf{det}(\mathbf{E}_t)|^{1/n}$. Then, we terminate the algorithm at any iteration $t \leq M$ when

$$\frac{\rho_t}{r} < \frac{\omega}{V_t + \omega}.$$

In our numerical experiments $M = 521$ whereas the average number of iterations $\widehat{M}$ with the new stopping rule was approximately $\widehat{M} = 464$. This reduction in the number of iterations significantly reduces the number of samples generated. In Table 1, the column labeled $N_{asc}(\beta)$ lists the number of samples required per iteration of SolveAmbChance for a given $\beta > 0$ ($N_{asc}(0)$ corresponds to SolveChance), the columns labeled $M \times N_{asc}(\beta)$ and $\widehat{M} \times N_{asc}(\beta)$ list the total number of samples needed with the old and the new stopping rule, respectively, and the last column summarizes the decrease in the number of samples.

β	$N_{asc}(\beta)$	$M \times N_{asc}(\beta)$	$\widehat{M} \times N_{asc}(\beta)$	$(M - \widehat{M}) \times N_{asc}(\beta)$
0.0300	150	78150	69600	8550
0.0200	100	52100	46400	5700
0.0150	86	44806	39904	4902
0.0100	75	39075	34800	4275
0.0050	67	34907	31088	3819
0	60	31260	27840	3420

Table 1: Number of iterations and samples as a function of the stopping rule

Performance evaluation We assume that the distribution $\mathbb{Q}$ of the random demand $\mathbf{D}$ is uncertain and is only known to belong to the uncertainty set $\mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}_0, \mathbb{Q}) \leq \beta, \mathbb{Q}_0 \sim \mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})\}$. Since checking whether a measure $\mathbb{Q} \in \mathcal{Q}$ is hard, we test the performance of the solutions using measures that are "similar" to $\mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$. In Section 5.2 we discuss performance of the solutions when the test measure $\mathbb{Q}$ is another Normal distribution with different parameters and in Section 5.3 we discuss the results for Cauchy and Pareto distributions.

The performance of the solutions was estimated by out-of-sample testing. We generated 50 batches of 1000 samples each (i.e. a total of 50000 samples). The error of a solution $\mathbf{x}$ with respect to a batch k, $k = 1, \ldots, 50$, is given by

$$\text{err}_k(\mathbf{x}) = \frac{1}{1000} \sum_{j=1}^{1000} (1 - \mathbf{1}_{\mathcal{R}(\mathbf{H}_j)}(\mathbf{x}))$$

where $\mathbf{1}_{\mathcal{A}}$ is the indicator function of set $\mathcal{A}$ and $\mathcal{R}(\mathbf{H}_j) = \{\mathbf{x} \mid \exists \mathbf{v} \in \mathbf{R}^k \text{ s.t. } \mathbf{W}\mathbf{v} \geq f(\mathbf{x}, \mathbf{H}_j)\}$. The error of $\mathbf{x}$, $\text{err}(\mathbf{x})$, is defined as the average of $\text{err}_k(\mathbf{x})$ over the 50 batches.

Problem parameters The mean demand vector $\bar{\mathbf{d}}$ was set to $\bar{\mathbf{d}} = (3, 2, 2)^T$ and standard deviation σ was set to $\sigma = 0.5 \times 10^{-2}$. The unit capacity costs are taken as $c_{s1} = 2$, $c_{s2} = 0.5$, $c_{s3} = 2$, and $c_{13} = 0.1$. The target violation probability ϵ was set to $\epsilon = 0.05$. The other parameters $\delta = 0.05$, $r = 10^{-2}$, and $\omega = 10^{-2}$.

5.2 Experiments with the Normal distribution

In this section we report the performances of the three solutions when the test samples are drawn from $\mathcal{N}(\bar{\mathbf{d}} + \widehat{\beta}\frac{\mathbf{e}}{\|\mathbf{e}\|}, \sigma^2\mathbf{I})$, where $\mathbf{e}$ is the vector of all ones and $\widehat{\beta} = 0.0475$, i.e. a Normal distribution with a shifted mean. As indicated in the previous section, the training distribution $\mathbb{Q}_0 = \mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$. The solutions $\mathbf{x}_{sc}$ and $\mathbf{x}_{asc}$ were computed from one run of Algorithms SolveChance and SolveAmbChance , respectively. We will comment on this later in this section.

Table 2 displays the performance of the three solutions strategies as a function of β. The columns labeled "err" list the empirical estimate of the violation probability, the column labeled

	$\mathbf{x}_{asc}$			$\mathbf{x}_{sc}$			$\mathbf{x}_{det}$		
$\widehat{\beta}$	err	stdev	c	err	stdev	c	err	stdev	c
0.0475	0	0	11.3888	**0.2131**	0.0019	11.1810	**1.0000**	0.0000	11.0000
0.0450	0	0	11.3790	**0.2623**	0.0014	11.1814	**0.9999**	0.0001	11.0000
0.0400	0.0000	0.0000	11.3277	**0.0955**	0.0014	11.1873	**0.9997**	0.0001	11.0000
0.0300	0.0000	0.0000	11.2912	**0.0707**	0.0014	11.1707	**0.9983**	0.0002	11.0000
0.0250	0.0000	0.0000	11.2798	0.0137	0.0006	11.1910	**0.9954**	0.0003	11.0000
0.0200	0.0001	0.0001	11.2509	0.0130	0.0005	11.1799	**0.9894**	0.0005	11.0000
0.0150	0.0001	0.0001	11.2358	0.0087	0.0003	11.2074	**0.9775**	0.0006	11.0000
0.0100	0.0002	0.0001	11.2162	0.0045	0.0003	11.1834	**0.9577**	0.0010	11.0000
0.0050	0.0005	0.0001	11.1922	0.0015	0.0002	11.1790	**0.9233**	0.0010	11.0000
0.0025	0.0003	0.0001	11.1916	0.0010	0.0001	11.1875	**0.9033**	0.0013	11.0000
0.0005	0.0005	0.0001	11.1759	0.0027	0.0002	11.1680	**0.8782**	0.0014	11.0000

Table 2: Performance when the test distribution $\mathbb{Q} = \mathcal{N}(\bar{\mathbf{d}} + \widehat{\beta}\frac{\mathbf{e}}{\|\mathbf{e}\|}, \sigma^2\mathbf{I})$.

"stdev" lists the standard deviation of the empirical estimate, and the column labeled "c" lists the cost of the solution. We highlight the instances where empirical estimate "err" violates the bound $\epsilon = 0.05$ by listing the value in bold face. The constant β defining the uncertainty set $\mathcal{Q}$ (see (22)) for the ambiguous chance constraint problem was set equal to $\widehat{\beta}$. We can draw the following conclusions from the results displayed in Table 2.

(i) The deterministic solution $\mathbf{x}_{det}$ does not have any tolerance for variance in demand. This is not surprising since it completely ignores the distributional information.

(ii) The chance constrained solution $\mathbf{x}_{sc}$ has more tolerance for random variations in demand as well as changes in the underlying distribution – the chance constrained solution meets bound on the violation probability $\epsilon = 0.05$ for all values of $\beta \leq 0.025$. However, this tolerance come with a higher capacity cost. As β is increased beyond the threshold value of 0.025, the performance of the chance constrained solution deteriorates very sharply.

(iii) The ambiguous chance constrained solution never violates the bound $\epsilon = 0.05$. In fact, setting the constant β defining $\mathcal{Q}$ equal to $\widehat{\beta}$ of the test distribution results in very conservative solutions – the violation probability estimates are considerably smaller than the allowable bound of ϵ. This also results in a correspondingly higher capacity cost.

Since setting $\beta = \widehat{\beta}$ resulted in very conservative solutions, we next investigated whether setting $\beta << \widehat{\beta}$ results in feasible solutions with costs comparable to that of the chance constrained solution $\mathbf{x}_{sc}$. Table 3 displays the performance of the ambiguous chance constrained solutions for different values of β when $\widehat{\beta} = 0.0475$. We see that $\mathbf{x}_{asc}(\beta)$ for all $\beta \geq 0.015$ is feasible with respect to $\widehat{\beta}$. Moreover, the cost of $\mathbf{x}_{asc}(0.015)$ is comparable to that of the chance constrained solution

	$\mathbf{x}_{asc}(\beta)$		
β	$\text{err}(\mathbf{x}_{asc}(\beta))$	$\text{stdev}(\mathbf{x}_{asc}(\beta))$	$c(\mathbf{x}_{asc}(\beta))$
0.0475	0	0	11.3799
0.0450	0	0	11.3589
0.0400	0	0	11.3393
0.0300	0.0020	0.0002	11.2867
0.0250	0.0015	0.0002	11.2935
0.0200	0.0078	0.0004	11.2570
0.0150	0.0184	0.0006	11.2406
0.0100	**0.0632**	0.0013	11.2154
0.0050	**0.0776**	0.0011	11.2109
0.0025	**0.2176**	0.0018	11.1806
0.0005	**0.2319**	0.0018	11.1790

Table 3: Performance when the test distribution $\mathbb{Q} = \mathcal{N}\big(\bar{\mathbf{d}} + \widehat{\beta}\frac{\mathbf{e}}{\|\mathbf{e}\|}, \sigma^2\mathbf{I}\big)$, with $\widehat{\beta} = 0.0475$.

$\widehat{\mathbf{x}}_{sc}$		
err	stdev	c
0.1853	0.0017	11.1851

	$\widehat{\mathbf{x}}_{asc}(\beta)$		
β	err	stdev	c
0.0300	0.0005	0.0001	11.3006
0.0200	0.0071	0.0004	11.2599
0.0150	0.0220	0.0007	11.2413
0.0100	0.0462	0.0010	11.2223
0.0050	**0.1150**	0.0015	11.1997

Table 4: Performances of $\widehat{\mathbf{x}}_{asc}(\beta)$ and $\widehat{\mathbf{x}}_{ac}$ when $\mathbb{Q} = \mathcal{N}(\bar{\mathbf{d}} + \widehat{\beta}\frac{\mathbf{e}}{\|\mathbf{e}\|}, \sigma^2\mathbf{I})$, $\widehat{\beta} = 0.0475$.

$\mathbf{x}_{sc}$ while at the same time providing insurance against all test distributions with $0 \leq \widehat{\beta} \leq 0.0475$. These results reiterate that $\mathbf{x}_{asc}(\beta)$ is very conservative for a given level β; therefore, the designer has to carefully select β to protect against over-designing.

The chance constraint is a non-convex constraint and the output of Algorithms SolveChance and SolveAmbChance are both random. Erdoğan and Iyengar [12] show that taking a convex combination of samples of a random solution significantly improves the performance in such a situation. We test this result by studying the performance of the solutions $\widehat{\mathbf{x}}_{sc}$ and $\widehat{\mathbf{x}}_{asc}$ that are generated by taking the empirical average over $N = 50$ independent runs. Table 4 presents the results related to $\widehat{\mathbf{x}}_{asc}(\beta)$ and $\widehat{\mathbf{x}}_{sc}$. It is clear from the results in Table 4 that averaging improves the performance. In particular, the $\widehat{\mathbf{x}}_{asc}(\beta)$ is feasible for $\widehat{\beta} = 0.0475$ even when one sets $\beta = 0.01$.

$\mathbf{x}_{sc}$		
err	stdev	c
0.0681	0.0081	11.1930

$\mathbf{x}_{det}$		
err	stdev	c
0.8722	0.0109	11.0000

	$\mathbf{x}_{asc}(\beta)$		
β	err	stdev	c
0.0475	0.0286	0.0044	11.3710
0.0400	0.0378	0.0058	11.3384
0.0300	0.0414	0.0067	11.2827
0.0250	0.0395	0.0057	11.2688
0.0200	0.0495	0.0063	11.2482
0.0150	0.0489	0.0069	11.2445
0.0100	0.0481	0.0066	11.2238
0.0050	**0.0631**	0.0064	11.2115
0.0025	**0.0712**	0.0074	11.1998

Table 5: Performance when test distribution $\mathbb{Q}$ is multivariate Cauchy with median $\bar{\mathbf{d}}$.

5.3 Experiments with other distributions

Table 5 displays the performance of the three solution strategies when the test distribution $\mathbb{Q}$ is a multivariate Cauchy distribution with median $\bar{\mathbf{d}}$, i.e. each $\mathbf{D}_i$ is a Cauchy random variable with median $\mathbf{d}_i$, $i = 1, 2, 3$. We assumed that $\mathbf{D}_i$, $i = 1, 2, 3$ are independent. Once again we see that the deterministic solution $\mathbf{x}_{det}$ and the chance constrained solution $\mathbf{x}_{sc}$ are both infeasible. The ambiguous chance constrained solution is feasible for all values of $\beta \geq 0.01$ and the additional cost for the robustness offered by the ambiguous chance constrained solution is at most 1.6%. Note that we have no guarantee that the Cauchy distribution belongs to the uncertainty set $\mathcal{Q}$ for any value of β.

Table 6 displays the performances of the three solutions when the test distribution is Pareto with variance $\sigma^2\mathbf{I}$ and a new mean $\widehat{\mathbf{d}}$ such that CDF of $\mathcal{N}(\bar{\mathbf{d}}, \sigma^2\mathbf{I})$ is greater than the CDF of the Pareto for all $\mathbf{d}$, i.e. each component of the chosen Pareto distribution stochastically dominates $\mathcal{N}(\bar{\mathbf{d}}_i, \sigma^2)$ for $i = 1, 2, 3$. The performance of the three solutions is very similar to the case when the test distribution was Cauchy – the ambiguous chance constrained solution provides robustness to perturbation in the distribution at a modest increase in cost.

6 Conclusion

In this paper we study two-stage convex chance constrained problems. Nemirovski and Shapiro [18] formulated this class of problems and proposed an ellipsoid-like iterative solution algorithm for the special case where the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ (see (7)) is bi-affine. We show that the Algorithm SolveChance (see Figure 1) extends the results in [18] to bi-convex $\mathbf{f}(\mathbf{x}, \mathbf{h})$ in a fairly straightforward fashion. The computational complexity of SolveChance as well as the quality of its output $\widehat{\mathbf{x}}$ depend on the radius r of the largest Euclidean ball that can be inscribed in the random set $\mathcal{X}_{\mathcal{I}}$ (see (11)) that is defined by the random set of linear inequalities generated during one run

$\mathbf{x}_{sc}$		
err	stdev	c
0.0554	0.0076	11.1846

$\mathbf{x}_{det}$		
err	stdev	c
0.9809	0.0045	11.0000

	$\mathbf{x}_{asc}(\beta)$		
β	err	stdev	c
0.0475	0.0013	0.0011	11.3842
0.0450	0.0016	0.0014	11.3626
0.0400	0.0036	0.0016	11.3297
0.0300	0.0080	0.0030	11.3004
0.0250	0.0082	0.0028	11.2753
0.0200	0.0108	0.0032	11.2443
0.0150	0.0241	0.0053	11.2402
0.0100	0.0255	0.0046	11.2230
0.0050	**0.0513**	0.0067	11.2221
0.0025	**0.0513**	0.0070	11.1911
0.0005	**0.0666**	0.0073	11.2259

Table 6: Performance when $\mathbb{Q}$ is multivariate Pareto with a shifted mean and variance $\sigma^2\mathbf{I}$.

of SolveChance. Since the set $\mathcal{X}_{\mathcal{I}}$ is random, selecting r is difficult; yet SolveChance requires r as an input. In this paper we provide some guidance for selecting r. We show that the largest value of r is related to the degree of robustness of the two-stage chance constrained problem – the more robust the problem, the higher one can set the parameter r. This is reminiscent of results relating the condition number of optimization problems to their computational complexity [20, 21, 13].

Next, we formulate ambiguous two-stage chance constrained problems. In this formulation, the random parameter $\mathbf{H}$ is known to have a fixed distribution $\mathbb{Q}$. However, the decision maker is only able to estimate that $\mathbb{Q}$ belongs to an uncertainty set $\mathcal{Q}$ of the form $\mathcal{Q} = \{\mathbb{Q} : \rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \beta\}$, where ρ_p denotes the Prohorov metric and β is an exogenously defined constant. We construct an algorithm SolveAmbChance that solves the ambiguous two-stage chance constrained problem when the impact function $\mathbf{f}(\mathbf{x}, \mathbf{h})$ is bi-affine and the extreme points of the dual polytope corresponding to the set $\mathcal{C}$ are explicitly known. The bi-affine assumption can be defended on the grounds that it still allows one to model a wide variety of applications; however, the latter assumption is much more serious and cannot be relaxed in general. We are currently exploring the possibility of replacing the adjustable robust characterization (24) by a chance constrained characterization. Let $\mathcal{L} = \{\boldsymbol{\lambda} : \mathbf{W}^T\boldsymbol{\lambda} = \mathbf{0}, \mathbf{1}^T\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \mathbf{0}\}$ and let $\mathbb{P}$ denote any probability measure on $\mathcal{L}$. Let $g(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}) = (\mathbf{a}_0(\mathbf{x}) + \mathbf{A}_1(\mathbf{x})\mathbf{h})^T\boldsymbol{\lambda} + \beta\|\mathbf{A}_1^T(\mathbf{x})\boldsymbol{\lambda}\|$. Then $\mathbf{x} \in \mathcal{R}_\beta$ if, and only if, $g(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}) \leq 0$, for all $\boldsymbol{\lambda} \in \mathcal{L}$, i.e. $\mathbb{P}(\boldsymbol{\lambda} : g(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}) \leq 0) = 1$. Currently, we are investigating how one can relax this constraint to $\mathbb{P}(\boldsymbol{\lambda} : g(\mathbf{x}, \mathbf{h}, \boldsymbol{\lambda}) \leq 0) > 1 - \epsilon$.

In the context of ambiguous chance constrained problems, the choice of the Prohorov metric can be justified by the fact that it is the natural metric for defining weak convergence of measures. It is, however, not clear how one should compute the constant β. We propose the following strategy. Suppose we assume that $\mathbf{H}$ is described by a parametric family of distributions $\mathcal{F}(\theta)$, $\theta \in \Theta$.

Suppose we estimate θ using an estimator $T_N(\mathbf{H}_1, \ldots, \mathbf{H}_N)$ and let $\mathbb{Q}_0 = \mathcal{F}(\theta_0)$, where $\theta_0 = T_N(\mathbf{H}_1, \ldots, \mathbf{H}_N)$. In the robust statistics literature, there is a "breakdown point" $\epsilon(T)$ associated with every estimator beyond which the estimator is completely unreliable [14]. Heuristically, the estimator is said to perform well for all measures $\mathbb{Q}$ such that $\rho_p(\mathbb{Q}, \mathbb{Q}_0) \leq \frac{1}{4}\epsilon(T)$. Thus, we could set $\beta = \frac{1}{4}\epsilon(T)$.

References

[1] M. Anthony and N. Biggs. *Computational Learning Theory.* Cambridge University Press, 1992.

[2] A. Atamtürk and M. Zhang. Two-stage robust network flow and design for demand uncertainty. Manuscript, 2005.

[3] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Math. Program.*, 99(2, Ser. A):351–376, 2004.

[4] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23(4):769–805, 1998.

[5] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization.* SIAM, Philadelphia, PA, 2001.

[6] D. Bertsimas and M. Sim. Robust conic optimization. Under review in *Math. Prog.*, May 2004.

[7] G. Calafiore and M. C. Campi. Uncertain convex programs: Randomized solutions and confidence levels. To appear in *Math. Prog.*, 2003.

[8] G. Calafiore and M. C. Campi. Decision making in an uncertain environment: the scenario-based optimization approach. Working paper, 2004.

[9] D. P. de Farias and B. Van Roy. On constraint sampling in the linear programming approach to approximate dynamic programming. To appear in *Math. Oper. Res.*, 2001.

[10] J. Dupačová. Stochastic programming: minimax approach. In *Encyclopedia of Optimization.* Kluwer, 2001.

[11] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. To appear in *Math. Prog.*, 2004.

[12] E. Erdoğan and G. Iyengar. Boosting, importance sampling, and convex chance constrained problem. Manuscript, 2005.

[13] R. M. Freund and J. R. Vera. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM J. Opt.*, 10:155–176, 1999.

[14] F. R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics: The approach based on influence functions.* John Wiley & Sons, 1986.

[15] M. J. Kearns and U. V. Vazirani. *An introduction to computational learning theory.* MIT Press, Cambridge, MA, 1997.

[16] L. G. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademiia Nauk SSSR*, 244(S):1093–1096, 1979. Translated in Soviet Mathematics Doklady 20:1 (1979).

[17] A. Nemirovski. On tractable approximations of randomly perturbed convex constraints. In *Proc. 42nd IEEE Conf. Dec. Contr. (CDC)*, volume 3, pages 2419– 2422, 2003.

[18] A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. To appear in *Probabilistic and randomized methods for design under uncertainty*, 2004.

[19] A. Prekopa. *Stochastic Programming.* Kluwer Academic Publishers, 1995.

[20] J. Renegar. Some perturbation theory for linear programming. *Math. Prog.*, 65:73–91, 1994.

[21] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Math. Prog.*, 70:279–351, 1995.

[22] A. Ruszczynski and A. Shapiro, editors. *Stochastic Programming.* Handbook in Operations Research and Management Science. Elsevier, 2003.

[23] A. Shapiro. Some recent developments in stochastic programming. *ORB Newsletter*, 13, March 2004. Available at http://www.ballarat.edu.au/ard/itms/CIAO/ORBNewsletter/issue13.shtml#11.

[24] A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. To appear in *SIAM J. Opt.*, 2004.

[25] A. Shapiro and A. J. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17:523–542, 2002.

[26] N. Z. Shor. Cut-off method with space extension in convex programming problems. *Cybernetics*, 13:94–96, 1977.

[27] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer, New York, NY, 1995.

[28] J. Žáčková. On minimax solutions of stochastic linear programs. *Čas. Pěst. Mat.*, pages 423–430, 1966.